

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号  
特許第7286268号  
(P7286268)

(45)発行日 令和5年6月5日(2023.6.5)

(24)登録日 令和5年5月26日(2023.5.26)

(51)国際特許分類		F I	
G 0 6 T	7/00 (2017.01)	G 0 6 T	7/00 3 5 0 C
G 0 6 T	7/60 (2017.01)	G 0 6 T	7/60 1 5 0 Z
G 0 1 B	11/26 (2006.01)	G 0 1 B	11/26 H

請求項の数 16 (全17頁)

(21)出願番号	特願2018-25033(P2018-25033)	(73)特許権者	000001007 キヤノン株式会社 東京都大田区下丸子3丁目30番2号
(22)出願日	平成30年2月15日(2018.2.15)	(74)代理人	100110412 弁理士 藤元 亮輔
(65)公開番号	特開2019-139694(P2019-139694 A)	(74)代理人	100104628 弁理士 水本 敦也
(43)公開日	令和1年8月22日(2019.8.22)	(74)代理人	100121614 弁理士 平山 倫也
審査請求日	令和3年2月1日(2021.2.1)	(72)発明者	井田 義明 東京都大田区下丸子3丁目30番2号 キヤノン株式会社内
前置審査		審査官	笠田 和宏

最終頁に続く

(54)【発明の名称】 画像処理方法、画像処理装置、撮像装置、画像処理プログラム、および、記憶媒体

(57)【特許請求の範囲】

【請求項1】

異なる光源環境で同一の被写体を撮影して得られた複数の撮影画像に基づき、前記異なる光源環境に対応する複数のチャンネルを有する入力データを取得する工程と、

法線情報に関する学習データを取得する工程と、

前記入力データと前記学習データとに基づいて前記被写体に関する法線情報を推定する工程と、を有し、

前記法線情報を推定する工程は、Nを2以上の整数、nを1からNまでの整数とするとき、

前記入力データに対して、前記学習データに基づくチャンネルの数を表す次元を含む3次元フィルタを用いた処理により、複数の線型関数のそれぞれによる第n線型変換と、非線型関数による第n非線型変換とをnが1からNになるまで順に実行することで中間データを生成する工程と、

前記中間データに対して、前記学習データに基づく少なくとも一つの線型関数による第N+1線型変換を実行する工程とを含むことを特徴とする画像処理方法。

【請求項2】

前記光源環境に関する光源情報を取得する工程を更に有し、

前記入力データのサイズは、前記光源情報に基づいて決定されることを特徴とする請求項1に記載の画像処理方法。

【請求項3】

前記被写体の位置情報を取得する工程を更に有し、

前記入力データのサイズは、前記位置情報に基づいて決定されることを特徴とする請求項 1 または 2 に記載の画像処理方法。

【請求項 4】

前記第  $n$  線型変換 ( $n = 1 \sim N$ ) のそれぞれは、前記学習データに基づく複数のフィルタの各々とのコンボリューションを含むことを特徴とする請求項 1 乃至 3 のいずれか 1 項に記載の画像処理方法。

【請求項 5】

前記法線情報を推定する工程において、前記第  $N + 1$  線型変換は、前記学習データに基づくフィルタとのコンボリューションを含むことを特徴とする請求項 4 に記載の画像処理方法。

10

【請求項 6】

前記光源環境に関する光源情報を取得する工程を更に有し、

前記第  $n$  線型変換 ( $n = 1 \sim N$ ) および前記第  $N + 1$  線型変換のそれぞれにおける前記フィルタのサイズは、前記光源情報に基づいて決定されることを特徴とする請求項 5 に記載の画像処理方法。

【請求項 7】

前記被写体の位置情報を取得する工程を更に有し、

前記第  $n$  線型変換 ( $n = 1 \sim N$ ) および前記第  $N + 1$  線型変換のそれぞれにおける前記フィルタのサイズは、前記位置情報に基づいて決定されることを特徴とする請求項 5 に記載の画像処理方法。

20

【請求項 8】

前記入力データは、前記被写体の距離マップに対応するチャンネルを有することを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の画像処理方法。

【請求項 9】

前記法線情報を推定する工程は、前記被写体の距離マップを推定することを特徴とする請求項 1 乃至 7 のいずれか 1 項に記載の画像処理方法。

【請求項 10】

前記入力データは、前記被写体の反射特性マップに対応するチャンネルを有することを特徴とする請求項 1 乃至 9 のいずれか 1 項に記載の画像処理方法。

30

【請求項 11】

前記法線情報を推定する工程は、前記被写体の反射特性マップを推定することを特徴とする請求項 1 乃至 9 のいずれか 1 項に記載の画像処理方法。

【請求項 12】

前記入力データは、鏡面反射成分が除去された画像に対応するチャンネルを有することを特徴とする請求項 1 乃至 11 のいずれか 1 項に記載の画像処理方法。

【請求項 13】

法線情報に関する学習データを記憶する記憶部と、

異なる光源環境で同一の被写体を撮影して得られた複数の撮影画像に基づく、前記異なる光源環境に対応する複数のチャンネルを有する入力データと、前記学習データとに基づいて前記被写体に関する法線情報を推定する画像処理部と、を有し、

前記画像処理部は、 $N$  を 2 以上の整数、 $n$  を 1 から  $N$  までの整数とするとき、

前記入力データに対して、前記学習データに基づくチャンネルの数を表す次元を含む 3 次元フィルタを用いた処理により、複数の線型関数のそれぞれによる第  $n$  線型変換と、非線型関数による第  $n$  非線型変換とを  $n$  が 1 から  $N$  になるまで順に実行することで中間データを生成し、

40

前記中間データに対して、前記学習データに基づく少なくとも一つの線型関数による第  $N + 1$  線型変換を実行することを特徴とする画像処理装置。

【請求項 14】

被写体空間の像を取得する撮像部と、

50

請求項 1 3 に記載の画像処理装置と、を有することを特徴とする撮像装置。

【請求項 1 5】

請求項 1 乃至 1 2 のいずれか 1 項に記載の画像処理方法をコンピュータに実行させることを特徴とする画像処理プログラム。

【請求項 1 6】

請求項 1 5 に記載の画像処理プログラムを記憶していることを特徴とする記憶媒体。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、被写体の法線情報を取得する画像処理方法に関する。

10

【背景技術】

【0002】

特許文献 1 には、4 つ以上の光源を使用して得られた複数の法線候補から真の面法線を求める方法が開示されている。非特許文献 1 には、畳み込みニューラルネットワークを応用して 1 枚の撮影画像から法線情報を推定する方法が開示されている。

【先行技術文献】

【特許文献】

【0003】

【文献】特開 2010 - 122158 号公報

【非特許文献】

20

【0004】

【文献】D. Eigen, et al. "Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture", arXiv:1411.4734 (2014).

Y. LeCun, et al., "Gradient-based Learning Applied to Document Recognition", Proc. of The IEEE, 1998.

G. E. Hinton, et al., "A fast learning algorithm for deep belief nets", Neural Comput. 2006 Jul; 18(7): 1527-54.

30

I. J. Goodfellow, et al., "Maxout networks", arXiv preprint arXiv:1302.4389 (2013).

G. E. Hinton & R. R. Salakhutdinov (2006-07-28). "Reducing the Dimensionality of Data with Neural Networks", Science 313(5786): 504-507.

N. Srivastava, et al., "Dropout: A simple way to prevent neural networks from overfitting", The Journal of Machine Learning Research, 15(1):1929-1958, 2014.

40

【発明の概要】

【発明が解決しようとする課題】

【0005】

しかし特許文献 1 の方法では、鏡面反射成分の影響を受ける撮影画像が複数ある場合、陰影部が生じた場合、ランバートの余弦則からずれた反射特性を持つ被写体の場合、相互反射が生じた場合、または、金属や透明体の場合等に、被写体の法線情報を推定できない。非特許文献 1 に開示された手法では、ある光源条件下での撮影画像のみから推定するため、被写体の反射率の変化と形状の影響を分離することが難しく、推定精度が十分でない。

【0006】

50

そこで本発明は、撮像画像から高精度に法線情報を推定することが可能な画像処理方法、画像処理装置、撮像装置、画像処理プログラム、および、記憶媒体を提供することを目的とする。

【課題を解決するための手段】

【0007】

本発明の一側面としての画像処理方法は、異なる光源環境で同一の被写体を撮影して得られた複数の撮影画像に基づく、前記異なる光源環境に対応する複数のチャンネルを有する入力データを取得する工程と、法線情報に関する学習データを取得する工程と、前記入力データと前記学習データとに基づいて前記被写体に関する法線情報を推定する工程とを有し、前記法線情報を推定する工程は、 $N$ を2以上の整数、 $n$ を1から $N$ までの整数とするとき、前記入力データに対して、前記学習データに基づくチャンネルの数を表す次元を含む3次元フィルタを用いた処理により、複数の線型関数のそれぞれによる第 $n$ 線型変換と、非線型関数による第 $n$ 非線型変換とを $n$ が1から $N$ になるまで順に実行することで中間データを生成する工程と、前記中間データに対して、前記学習データに基づく少なくとも一つの線型関数による第 $N + 1$ 線型変換を実行する工程とを含む。

10

【0008】

本発明の他の側面としての画像処理装置は、法線情報に関する学習データを記憶する記憶部と、異なる光源環境で同一の被写体を撮影して得られた複数の撮影画像に基づく、前記異なる光源環境に対応する複数のチャンネルを有する入力データと、前記学習データとに基づいて前記被写体に関する法線情報を推定する画像処理部とを有し、前記画像処理部は、 $N$ を2以上の整数、 $n$ を1から $N$ までの整数とするとき、前記入力データに対して、前記学習データに基づくチャンネルの数を表す次元を含む3次元フィルタを用いた処理により、複数の線型関数のそれぞれによる第 $n$ 線型変換と、非線型関数による第 $n$ 非線型変換とを $n$ が1から $N$ になるまで順に実行することで中間データを生成し、前記中間データに対して、前記学習データに基づく少なくとも一つの線型関数による第 $N + 1$ 線型変換を実行する。

20

【0009】

本発明の他の側面としての撮像装置は、被写体空間の像を取得する撮像部と前記画像処理装置とを有する。

【0010】

本発明の他の側面としての画像処理プログラムは、前記画像処理方法をコンピュータに実行させる。

30

【0011】

本発明の他の側面としての記憶媒体は、前記画像処理プログラムを記憶している。

【0012】

本発明の他の目的及び特徴は、以下の実施例において説明される。

【発明の効果】

【0013】

本発明によれば、撮像画像から高精度に法線情報を推定することが可能な画像処理方法、画像処理装置、撮像装置、画像処理プログラム、および、記憶媒体を提供することができる。

40

【図面の簡単な説明】

【0014】

【図1】実施例1および実施例2における法線情報を推定するネットワーク構造を示す図である。

【図2】実施例1における撮像装置のブロック図である。

【図3】実施例1における撮像装置の外観図である。

【図4】実施例1における法線情報の推定処理を示すフローチャートである。

【図5】実施例1における学習データの学習を示すフローチャートである。

【図6】実施例2における画像処理システムのブロック図である。

50

【図7】実施例2における画像処理システムの外觀図である。

【図8】実施例2における法線情報の推定処理を示すフローチャートである。

【発明を実施するための形態】

【0015】

以下、本発明の実施例について、図面を参照しながら詳細に説明する。各図において、同一の部材については同一の参照符号を付し、重複する説明は省略する。

【0016】

デジタルカメラ等の撮像装置で被写体を撮像して得られた撮影画像から、被写体の形状情報として面法線の情報（以下、法線情報という）を取得する方法が知られている。法線情報を取得する方法としては、Shape from Shadingや、照度差ステレオ法がある。Shape from Shadingは、1枚の撮影画像からも法線情報を推定できるが、対象物体の反射率が一樣であることや被写体の形状がなめらかに変化すること等の仮定を必要とする。照度差ステレオ法は、被写体の面法線と光源方向とに基づいて反射特性を仮定し、複数の光源位置での被写体の輝度情報と仮定した反射特性とから面法線を決定する方法である。複数の光源位置で撮像した撮影画像を用いることにより、Shape from Shadingよりも少ない仮定の下で法線情報を推定することができる。仮定される被写体の反射特性として、ランバートの余弦則に従うランバート反射モデルが用いられることが多い。

【0017】

一般に、物体での反射には、鏡面反射と拡散反射とがある。鏡面反射は、物体表面での正反射であり、物体表面（界面）においてフレネルの式に従うフレネル反射である。拡散反射は、被写体の表面を透過した後物体内部で散乱されて光が返ってくる反射である。鏡面反射した光は、前述のランバートの余弦則で表すことはできない。このため、撮像装置で観測される被写体からの反射光に鏡面反射光が含まれていると、Shape from Shadingや照度差ステレオ法では面法線を正確に求めることができない。また、光源からの光が当たらない陰影部においても、仮定した反射モデルからのずれが生じ、被写体の法線情報を正確に取得することができない。また、表面の粗い被写体や半透明体などに関しては、拡散反射成分もランバートの余弦則からずれを生じる。また、相互反射が生じている場合および拡散反射成分が観測されない金属や透明体などにおいても、被写体の法線情報を正確に取得することができない。

【0018】

本発明の各実施例は、ディープラーニング（深層学習とも呼ばれる）を用いて、撮影画像から被写体の法線情報を推定する。ここで法線情報とは、法線方向ベクトルや、法線を表す各自由度を指す。各実施例によれば、異なる光源環境下で被写体を撮影した複数の撮影画像とその法線情報とを用いて、対応関係をディープラーニングで学習することにより、法線情報を高精度に推定することができる。

【実施例1】

【0019】

まず、図2および図3を参照して、本発明の実施例1における撮像装置について説明する。図2は、撮像装置100のブロック図である。図3は、撮像装置100の外觀図である。本実施例において、撮像装置100は、画像処理方法を実行し、ディープラーニングを用いて撮影画像から被写体の法線情報を推定する。

【0020】

撮像装置100は、被写体空間の像を撮影画像（入力画像）として取得する撮像部101を有する。撮像部101は、被写体空間から入射する光を集光する結像光学系101aと、複数の画素を有する撮像素子101bとを有する。撮像素子101bは、例えば、CCD（Charge Coupled Device）センサやCMOS（Complementary Metal-Oxide Semiconductor）センサである。

【0021】

画像処理部102は、撮影画像（入力画像、部分画像）に対して、法線情報を推定する

。画像処理部102は、学習部102aと推定部102bとを有する。記憶部103は、画像処理部102による法線情報を推定する際に用いられる学習データ（法線情報に関して予め学習された学習データ）を記憶している。この処理の詳細については後述する。画像処理部102により推定された法線情報は、液晶ディスプレイなどの表示部104に表示されるか、または、記録媒体105に保存される。ただし、法線情報に代えて、法線情報に基づいて生成された画像（例えば、レンダリング画像）を表示部104に表示または記録媒体105に保存してもよい。また、撮影画像（部分画像）を記録媒体105に保存し、画像処理部102が任意のタイミングで法線情報を推定するように構成してもよい。光源110は、撮像時（撮影画像の取得時）に選択的に点灯する。このため撮像部101は、複数の異なる光源環境下で撮影を行うことができる。光源110は撮像装置100と一体的に構成されているが、これに限定されるものではない。光源110に代えて、撮像装置100から分離した外部装置を光源として用いてもよい。以上の一連の制御は、システムコントローラ106により行われる。

10

#### 【0022】

図3に示されるように、光源110は、複数の光源部110a～110hを備えて構成される。本実施例の撮像装置100は、複数の光源部として8つの光源部110a～110hを有するが、これに限定されるものではなく、複数の画像を取得するために少なくとも2つ以上の光源を備えていればよい。また、撮像光学系の光軸OAから等距離の位置に同心円状に8つの光源部110a～110hを等間隔で配置しているが、これに限定されるものではない。また、複数の光源部110a～110hは、撮像装置100に内蔵されているが、これに限定されるものではなく、撮像装置100に対して着脱可能に構成されていてもよい。

20

#### 【0023】

次に、図4を参照して、画像処理部102による法線情報の推定処理について説明する。画像処理部102は、法線情報の推定処理の際に、事前に学習された学習データを用いるが、この学習の詳細については後述する。図4は、法線情報の推定処理を示すフローチャートである。図4の各ステップは、主に、画像処理部102の推定部102bにより実行される。

#### 【0024】

まずステップS101において、画像処理部102（推定部102b）は、異なる光源環境下で被写体を撮影した複数の撮影画像（入力画像）と学習データとを取得する。学習データとは、撮影画像と推定された法線情報とを結び付けるために予め学習された情報（学習情報）である。続いてステップS102において、推定部102bは、撮影画像から入力データ（撮影画像の少なくとも一部の部分画像）を取得する。部分画像は、各光源環境下における撮影画像のそれぞれに対して、画像上の同一位置を切り出すことで取得される。なお、各光源環境下における撮影画像のそれぞれに対して、電子的な手振れ補正処理等の位置ずらし処理を行ってもよい。また、撮影した異なる複数の光源環境のうち少なくとも二つ以上の一部の光源環境下における撮影画像のみに関して部分画像を取得してもよい。法線情報の推定は、部分画像（入力データ）を単位として部分画像ごとに行われる。

30

#### 【0025】

続いてステップS103において、推定部102bは、学習データを用いて入力データから法線情報を推定した部分法線情報を生成する（部分法線情報を推定する）。ここで、図1を参照して、法線情報の推定処理について詳述する。図1は、ディープラーニングの一つであるCNN（Convolutional Neural Network）のネットワーク構造を示している。ただし、ディープラーニングとして、CNN以外の手法、例えばDBN（Deep Belief Network）を用いても構わない。CNNおよびDBNの詳細はそれぞれ、非特許文献2および非特許文献3に説明されている。

40

#### 【0026】

CNNは、複数の層構造になっており、各層で学習データを用いた線型変換と非線型変換とが実行される。nを1からNまでの整数とすると、n番目の層を第n層、第n層に

50

おける線型変換と非線型変換とをそれぞれ、第  $n$  線型変換と第  $n$  非線型変換と呼称する。ただし、 $N$  は 2 以上の整数である。部分画像 201 に関しては、第 1 層において、複数のフィルタ 202 のそれぞれとのコンボリューション（複数の線型関数による第 1 線型変換）が実行される。その後、活性化関数（Activation Function）と呼ばれる非線型関数を用いて変換（第 1 非線型変換）が実行される。図 1 において、活性化関数を  $AF$  として示している。また、部分画像 201 が複数枚描画されているのは、入力画像（撮影画像）が複数のチャンネルを有するためである。本実施例において、部分画像は RGB（Red、Green、Blue）の 3 チャンネルに対して、さらに異なる光源環境の分だけのチャンネルを有する。ただし、チャンネルの数はこれに限定されるものではない。また、部分画像が RGB のチャンネルを有する場合でも、光源環境についてのチャンネルは有するままで、各色で個別に CNN へ入力しても構わない。

10

## 【0027】

フィルタ 202 は複数存在する。推定部 102b は、複数のフィルタ 202 のそれぞれと部分画像 201 とのコンボリューションを個別に算出する。フィルタ 202 の係数は、学習データに基づいて決定される。学習データは、フィルタ 202 の係数（フィルタ係数）そのもの、または、フィルタ 202 を所定の関数でフィッティングした際の係数でもよい。フィルタ 202 のそれぞれのチャンネル数は、部分画像 201 の数と一致する。部分画像 201 のチャンネル数が 2 以上の場合、3 次元フィルタとなる（3 次元目がチャンネル数を表す）。また、コンボリューションの結果に対して、学習データから決定される定数（負もとり得る）を加算してもよい。

20

## 【0028】

活性化関数  $f(x)$  の例としては、以下の式 (1) ~ (3) が挙げられる。

## 【0029】

## 【数 1】

$$f(x) = \frac{1}{1 + e^{-x}} \quad \dots(1)$$

$$f(x) = \tanh x \quad \dots(2)$$

$$f(x) = \max(x, 0) \quad \dots(3)$$

30

## 【0030】

式 (1) はシグモイド関数、式 (2) はハイパボリックタンジェント関数、式 (3) は ReLU (Rectified Linear Unit) と呼ばれる。式 (3) 中の  $\max$  は、引数のうち最大値を出力する MAX 関数を表す。式 (1) ~ (3) に示される活性化関数  $f(x)$  は、全て単調増加関数である。また、活性化関数として Maxout を使用してもよい。Maxout は、第  $n$  線型変換の出力である複数の画像のうち、各画素で最大値である信号値を出力する MAX 関数である。Maxout の詳細は、非特許文献 4 に説明されている。

40

## 【0031】

図 1 において、第 1 線型変換および第 1 非線型変換が施された部分画像を、第 1 変換部分画像 203 と呼称する。第 1 変換部分画像 203 の各チャンネル成分は、部分画像 201 と複数のフィルタ 202 のそれぞれとのコンボリューションから生成される。このため、第 1 変換部分画像 203 のチャンネル数は、フィルタ 202 の数と同じになる。

## 【0032】

第 2 層では、第 1 変換部分画像 203 に対して、第 1 層と同様に学習データから決定される複数のフィルタ 204 とのコンボリューション（第 2 線型変換）と、活性化関数による非線型変換（第 2 非線型変換）とを行う。第 2 層で用いられるフィルタ 204 は、一般

50

的に、第1層で用いられるフィルタ202と同一ではない。フィルタ204のサイズや数も、フィルタ204と一致しなくてもよい。ただし、フィルタ204のチャンネル数と第1変換部分画像203のチャンネル数とは互いに一致する。推定部102bは、同様の演算を第N層まで繰り返す(第n線型変換および第n非線型変換( $n = 1 \sim N$ ))を実行することにより、中間データ210を取得する。

#### 【0033】

最後に、第N+1層において、中間データ210と複数のフィルタ211のそれぞれとのコンボリューションに定数を加算すること(第N+1線型変換)により、部分画像201に対する法線情報を推定した部分法線情報212が取得される。ここで用いられるフィルタ211および定数もそれぞれ、学習データに基づいて決定される。部分法線情報212のチャンネル数は、法線情報の表現方法に応じて異なる。法線ベクトルの3次元の成分を各チャンネルに割り当てれば3チャンネルであり、法線ベクトルの方向を2つの角度で表現すれば2チャンネルである。また、例えばRGBごとに個別に法線ベクトルを推定してもよく、法線情報の表現方法は限定されない。したがって、フィルタ211の数は部分法線情報212の数と同じになる。部分法線情報212の各チャンネルの成分は、中間データ210とフィルタ211のそれぞれとのコンボリューションを含む演算から求められる。なお、部分画像201と部分法線情報212とのサイズは互いに一致しなくてもよい。コンボリューションの際に、部分画像201の外側にはデータが存在しないため、データの存在する領域のみで演算すると、コンボリューション結果はサイズが小さくなる。ただし、周期境界条件などを設定することにより、サイズを保つこともできる。なお本実施例において、第m線型変換( $m = 1 \sim N + 1$ )のそれぞれに関する各フィルタの係数は、全て異なっている。

#### 【0034】

ディープラーニングが高い性能を発揮できる理由は、非線型変換を多層構造によって何度も実行することにより、高い非線型性が得られるためである。仮に、非線型変換を担う活性化関数が存在せず、線型変換のみでネットワークが構成されていた場合、いくら多層にしてもそれと等価な単層の線型変換が存在するため、多層構造にする意味がない。ディープラーニングは、より多層にする方が強い非線型を得られるため、高い性能が出やすいと言われている。一般に、少なくとも3層以上を有する場合がディープラーニングと呼ばれる。

#### 【0035】

続いて、図4のステップS104において、推定部102bは、撮影画像のうち所定の領域の全てに対して法線情報の推定処理が完了したか否かを判定する。所定の領域(部分画像)の全てに対して部分法線情報212が生成されている場合、ステップS105へ進む。一方、法線情報の推定処理が完了していない領域(部分画像)が残っている場合、ステップS102へ戻り、推定部102bは、まだ法線情報が推定されていない部分画像を撮影画像から取得する。

#### 【0036】

ステップS105において、推定部102bは、法線情報を出力する。法線情報は、生成された複数の部分法線情報212を合成することにより生成される。ただし、部分画像が撮影画像の全体である場合、推定部102bは、部分法線情報212をそのまま法線情報として出力する。以上の処理により、撮影画像から推定した法線情報を取得することができる。

#### 【0037】

なお本実施例において、視点の異なる複数の撮影画像を入力し、法線情報を取得するように構成してもよい。視点によっても光の反射角が変化することから被写体の反射特性の影響も変化する。同一の被写体に対して異なる光の反射を受けた複数の画像を入力することにより、推定精度を向上することができる。

#### 【0038】

次に、図5を参照して、本実施例における学習データの学習について説明する。図5は

10

20

30

40

50

、学習データの学習を示すフローチャートである。図5の各ステップは、主に、画像処理部102の学習部102aにより行われる。ただし本実施例はこれに限定されるものではなく、学習データの学習は、法線情報の推定前であれば、撮像装置100とは別の装置（演算装置）に設けられた学習部で行ってもよい。本実施例では、撮像装置100の学習部102aが学習データを学習する場合について説明する。

【0039】

まずステップS201において、学習部102aは、少なくとも一対の学習データを取得する。一対の学習データとは、異なる光源環境下で被写体を撮影した複数の撮影画像と、同一の被写体の法線情報である。1つの法線情報に対して対応させる複数の撮影画像の組合せは、光源環境の組合せをある1つのものだけ対応させてもよいし、複数の組合せを対応させてもよい。後者は、同じ法線情報を出力するために入力する複数撮影画像として、光源環境の組合せが1通りでなくともよいことに相当する。

10

【0040】

学習データとして用いる法線情報には、様々な形状や法線ベクトルを有する被写体、および、様々な反射特性の被写体が含まれていることが好ましい。例えば、学習データに光沢のある被写体がない場合、光沢の変化から法線を推定する学習データが存在しないため、光沢のある被写体に対する法線情報の推定の効果が十分に得られない可能性がある。

【0041】

学習データを用意する方法として、シミュレーションを利用してもよいし、実測した情報を使用してもよい。シミュレーションを行なう場合は、反射特性を付与した3Dモデルに対して、様々な光源条件でのCGレンダリングを行うことで異なる光源環境下での撮影画像に相当する画像を生成すればよい。

20

【0042】

実測した情報を使用する場合、既知の形状の被写体（法線情報が既知の被写体）を異なる光源環境下で撮影することで、複数の異なる光源環境下で撮影した撮影画像と対応する法線情報を得てもよい。

【0043】

また、被写体の撮影画像に対してディープラーニング以外の手法を用いて取得された法線情報を用いて、学習データを用意してもよい。その例として、レーザ測距により形状を取得する方法、構造化照明で照明した画像から形状を取得する方法、鏡面反射の反射方向と光源の入射方向から形状を取得する方法、または、接触式の形状計測機器で取得する方法がある。被写体のサイズ、形状、および、反射特性などの条件に応じて適切な取得方法を選択することにより、一般の被写体に対して高精度に法線情報を取得することができる。

30

【0044】

続いてステップS202において、学習部102aは、ステップS201にて取得した学習データから、複数の学習ペアを取得する。学習ペアは、学習用部分画像と学習用部分法線情報とからなる。学習用部分画像は複数の光源環境下で撮影された撮影画像から取得され、そのサイズはステップS102にて取得した部分画像と同じである。学習用部分法線情報は、法線情報から、その中心が学習用部分画像と同じ被写体位置になるように取得される。そのサイズは、ステップS103にて生成された部分法線情報と同じである。前述と同様に、学習用部分画像と学習用部分法線情報のペア（学習ペア）は、一対一に対応している必要はない。一つの学習用部分法線情報と、複数の学習用部分画像とがペア（グループ）になっていてもよい。

40

【0045】

続いてステップS203において、学習部102aは、複数の学習ペア（学習用部分画像と学習用部分法線情報）から、学習データを学習によって取得（生成）する。学習では、法線情報を推定するネットワーク構造と同じネットワーク構造を使用する。本実施例では、図1に示されるネットワーク構造に対して学習用部分法線情報を入力し、その出力結果と学習用部分画像との誤差を算出する。法線情報の誤差は、各成分の差分をとってもよく、または、学習用部分法線情報の法線ベクトルと出力結果の法線ベクトルとの内積を1

50

から引いた値としてもよい。この誤差が最小となるように、例えば誤差逆伝播法 (Backpropagation) などを用いて、第1乃至N+1層で用いる複数のフィルタのそれぞれの係数や加算する定数 (学習データ) を更新して最適化する。各フィルタの係数および定数の初期値は任意に設定することができ、例えば乱数から決定される。または、各層ごとに初期値を事前学習するAuto Encoderなどのプレトレーニングを行ってもよい。Auto Encoderの詳細は、非特許文献5に説明されている。

#### 【0046】

学習ペアの全てをネットワーク構造へ入力し、それら全ての情報を使って学習データを更新する手法をバッチ学習と呼ぶ。ただし、この学習方法は、学習ペアの数が増えるにつれて計算負荷が膨大になる。逆に、学習データの更新に一つの学習ペアのみを使用し、更新ごとに異なる学習ペアを使用する学習手法をオンライン学習と呼ぶ。この手法は、学習ペアが増えても計算量が増大しないが、一つの学習ペアに存在するノイズの影響を大きく受ける。このため、これら2つの手法の中間に位置するミニバッチ法を用いて学習することが好ましい。ミニバッチ法は、全学習ペアの中から少数を抽出し、それらを用いて学習データの更新を行う。次の更新では、異なる小数の学習ペアを抽出して使用する。これを繰り返すことにより、バッチ学習とオンライン学習の不利な点を小さくすることができ、高い推定効果を得やすくなる。

#### 【0047】

続いてステップS204において、学習部102aは、学習された学習データを出力する。本実施例において、学習データは記憶部103に記憶される。以上の処理により、反射特性や形状の制約がない一般的な被写体において、撮像画像から高精度に法線情報を推定可能な学習データを学習することができる。

#### 【0048】

また、以上の処理に加えて、CNNの性能を向上させる工夫を併用してもよい。例えば、ロバスト性の向上のためネットワークの各層において、ドロップアウト (Dropout) やダウンサンプリングであるプーリング (pooling) を行ってもよい。ドロップアウトの詳細は非特許文献6に説明されている。

#### 【0049】

本実施例では撮影画像から部分画像を取得しているが、撮影画像に任意の画像処理を行った後の画像から部分画像を取得してもよい。例えば、撮影画像から鏡面反射の影響を受けない (鏡面反射成分が除去された) 拡散反射画像を公知の方法で生成し、拡散反射画像から部分画像を取得してもよい。これにより、反射特性の異なる被写体においても鏡面反射成分が除去された成分のみに対して学習および推定ができ、推定精度を向上することができる。

#### 【0050】

また、入力データとして、部分画像に加えて、被写体の距離マップを用いることができる。距離マップは、部分画像と対応する領域のみ用いればよい。この場合、入力データのチャンネル数が1だけ増える。同一の光源環境においても被写体の距離によって被写体に入射する光量や入射方向ベクトルが異なる。この情報を入力データとして用いて学習および推定を行うことにより、法線情報の推定精度を向上することができる。被写体の距離マップは、例えば異なる視点から撮像した互いに視差を有する複数の視差画像に基づいて取得してもよい。撮像画像に基づいて取得した距離マップを用いることで、異なる光源環境下における撮影画像との位置合わせなどを行う必要がなくなるため、好ましい。

#### 【0051】

また、入力データとして、部分画像に加えて、被写体の反射特性マップを用いることができる。反射特性マップは、部分画像と対応する領域のみ用いればよい。この場合、入力データのチャンネル数が反射特性のチャンネル数だけ増える。例えば、拡散反射成分の反射率 (アルベド) と、鏡面反射成分の反射率および反射ピークの広がりをを用いる場合、3チャンネル増える。撮影画像における被写体の輝度の違いは、被写体の法線情報と反射特性とに応じて異なる。被写体の反射特性が既知の場合または何らかの手段で取得可能であ

10

20

30

40

50

る場合、反射特性の情報を入力データとして用いることにより、輝度の変化要因を法線情報に限定することができる。このため、反射特性マップを入力データとして用いて学習および推定を行うことにより、法線情報の推定精度を向上することができる。

#### 【0052】

また、入力データとして、部分画像に加えて、光源情報を示すデータを用いることもできる。例えば、光源の空間強度分布を示す光源環境マップを入力すればよいが、これに限定されるものではなく、光源情報を示すデータであればよい。この場合、入力データのチャンネル数が光源情報のチャンネル数だけ増える。部分画像が撮影された光源環境ごとにRGBの光源環境マップを用いる場合、光源環境の数の3倍だけチャンネル数が増える。撮影画像における被写体の輝度の違いは、光源情報と被写体の法線情報と反射特性とに応じて異なる。既知の光源環境下で撮影した撮影画像を用いる場合、光源情報を入力データとして用いることにより、輝度の変化要因から光源情報の違いを除外することが可能である。このため、光源情報を入力データとして用いて学習および推定を行うことにより、法線情報の推定精度を向上することができる。

10

#### 【0053】

また、法線情報を推定する際に、さらに被写体の距離マップを推定することもできる。この場合、部分法線情報212を1チャンネル増やして部分距離マップを出力すればよい。同一の光源環境においても被写体の距離に応じて被写体に入射する光量や入射方向ベクトルが異なり、撮影画像の輝度も変化する。このため、法線情報と距離マップとは相関する情報であり、同時に推定することが可能である。法線情報に加えて被写体の距離マップを推定する場合、学習用データにも距離マップを加える必要がある。これにより、距離マップを反映したより精度の高い学習が可能であり、推定精度も向上する。

20

#### 【0054】

また、法線情報を推定する際に、さらに被写体の反射特性マップを推定することもできる。この場合、部分法線情報212を推定させる反射特性のチャンネル数だけ増やして部分反射特性マップを出力すればよい。撮影画像における被写体の輝度の違いは、被写体の法線情報と反射特性とに応じて異なる。このため、法線情報と反射特性とは相関する情報であり、同時に推定することが可能である。法線情報に加えて被写体の反射特性マップを推定する場合、学習用データにも反射特性マップを加える必要がある。これにより、反射特性マップを反映したより精度の高い学習が可能であり、推定精度も向上する。

30

#### 【0055】

また、法線情報を推定する際に、さらに光源情報を推定することもできる。例えば、光源の空間強度分布を示す光源環境マップを出力すればよいが、これに限定されるものではなく、光源情報を示すデータであればよい。撮影画像における被写体の輝度の違いは、光源情報と被写体の法線情報と反射特性とに応じて異なる。このため、法線情報と光源情報とは相関する情報であり、同時に推定することが可能である。法線情報に加えて光源情報を推定する場合、学習用データにも光源情報を加える必要がある。これにより、光源情報を反映したより精度の高い学習が可能であり、推定精度も向上する。

#### 【0056】

本実施例によれば、反射特性や形状の制約がない一般的な被写体において、撮像画像から高精度に法線情報を推定可能な画像処理装置および撮像装置を提供することができる。

40

#### 【実施例2】

#### 【0057】

次に、本発明の実施例2における画像処理システムについて説明する。本実施例の画像処理システムにおいて、法線情報を推定する画像処理装置と、複数の異なる光源環境下で複数の撮影画像を取得する撮像装置、および、学習を行うサーバが個別に設けられている。また本実施例では、光源環境に関する光源情報を取得することにより、使用する学習データを切り替える。各光源情報に対して推定に使用する学習データを個別に学習して使用することで、より高精度な推定が可能となる。

50

## 【 0 0 5 8 】

図 6 および図 7 を参照して、本実施例における画像処理システムについて説明する。図 6 は、画像処理システム 2 0 0 のブロック図である。図 7 は、画像処理システム 2 0 0 の外観図である。図 6 および図 7 に示されるように、画像処理システム 2 0 0 は、撮像装置 3 0 0、画像処理装置 3 0 1、サーバ 3 0 5、表示装置 3 0 8、記録媒体 3 0 9、および、出力装置 3 1 0 を備えて構成される。

## 【 0 0 5 9 】

撮像装置 3 0 0 の基本構成は、法線情報の推定と学習データの学習に関する画像処理部を除いて、図 2 を参照して説明した撮像装置 1 0 0 と同様である。撮像装置 3 0 0 を用いて撮影された撮影画像（入力画像）は、画像処理装置 3 0 1 に設けられた記憶部 3 0 2 に記憶される。画像処理装置 3 0 1 は、ネットワーク 3 0 4 と有線または無線で接続されており、ネットワーク 3 0 4 を介してサーバ 3 0 5 にアクセスすることができる。サーバ 3 0 5 は、複数の異なる光源環境下で撮影された撮影画像から法線情報を推定するための学習データを学習する学習部 3 0 7 と、学習データを記憶する記憶部 3 0 6 とを有する。画像処理装置 3 0 1 に設けられた推定部 3 0 3（画像処理部）は、サーバ 3 0 5 の記憶部 3 0 6 からネットワーク 3 0 4 を介して学習データを取得し、法線情報を推定する。推定部 3 0 3 により推定された法線情報は、表示装置 3 0 8、記録媒体 3 0 9、および、出力装置 3 1 0 の少なくとも一つに出力される。なお、法線情報に代えて、法線情報に基づいて生成される画像（例えばレンダリング画像）を出力してもよい。表示装置 3 0 8 は、例えば液晶ディスプレイやプロジェクタである。ユーザは、表示装置 3 0 8 を介して、処理途中の画像を確認しながら作業を行うことができる。記録媒体 3 0 9 は、例えば半導体メモリ、ハードディスク、ネットワーク上のサーバである。出力装置 3 1 0 は、例えばプリンタである。画像処理装置 3 0 1 は、必要に応じて現像処理やその他の画像処理を行う機能を有してもよい。

## 【 0 0 6 0 】

次に、図 8 を参照して、推定部 3 0 3 により実行される法線情報の推定処理について説明する。図 8 は、法線情報の推定処理を示すフローチャートである。図 8 の各ステップは、主に、画像処理装置 3 0 1 の推定部 3 0 3（画像処理部）により実行される。

## 【 0 0 6 1 】

まずステップ S 3 0 1 において、推定部 3 0 3 は、記憶部 3 0 2 から、複数の異なる光源環境下で撮影された撮影画像を取得する。続いてステップ S 3 0 2 において、推定部 3 0 3 は、記憶部 3 0 2 から撮影画像を撮影した光源環境に関する光源情報を取得する。各光源環境における光源情報は、撮像装置 3 0 0 で撮影画像を撮影する際に、撮影画像と紐付けて記憶しておけばよい。光源情報は、光源の位置、光源の強度、または、波長特性などの被写体が照明される環境を示す情報であればよく、光源の大きさ、配置、強度分布、または、配光分布などを含む光源の空間分布情報として取得してもよい。撮像装置 3 0 0 のように光源が撮像装置に対して固定される場合、どの光源が点灯したかに関する情報を取得してもよい。また、間接的に光源情報を取得する情報として、光源情報のかわりに、被写体の距離や方向などの被写体の位置情報を取得してもよい。前述の各光源環境に対する情報を、撮影画像を撮影した複数の光源環境に対して総合した情報を光源情報として取得すればよい。なお、撮影した光源環境のうち一部の光源環境下における撮影画像に対してのみ以下の工程を行う場合、用いる光源環境に関する光源情報のみを取得すればよい。被写体の距離情報は、例えば撮像部のフォーカス情報から取得してもよく、異なる視点から撮像した互いに視差を有する複数の視差画像に基づいて取得してもよい。被写体の方向は、画面上の位置から取得することができる。

## 【 0 0 6 2 】

続いてステップ S 3 0 3 において、推定部 3 0 3 は、ステップ S 3 0 2 にて取得された光源情報に基づいて、使用するネットワーク構造と学習データ、および、部分画像のサイズ（大きさ）を決定する。本実施例において、推定部 3 0 3 は、図 1 に示される CNN を利用して法線情報を推定する。また本実施例では、光源情報に応じて個別に学習させた学

習データを用いるため、ネットワーク構造及び部分画像のサイズも学習時に使用したものと同じになる。取得した光源情報に応じた学習データを用いることで、より高精度な法線情報推定が可能となる。異なる光源環境下における部分画像を入力するため、部分画像のサイズが小さくても光源環境による輝度の違いから法線情報を推定することができる。一方、部分画像のサイズをある程度大きくとることにより、例えば遠近法やテクスチャの密度などに基づく大局的な情報も取り入れた推定を行うことができ、一般的な被写体に対して法線情報推定を行う上で精度の向上が可能となる。例えば、撮影画像を取得した複数の異なる光源環境間での差異（光源位置や空間強度分布など）が小さい場合、光源環境による輝度の違いが小さくなり、推定精度が不足する場合がある。これを回避するため、光源環境間の差異が小さい場合には、部分画像のサイズを大きくしてより大局的な情報を用いることが好ましい。

10

**【0063】**

ネットワーク構造は、各層で使用するフィルタのサイズだけでなく、1つの層で使われるフィルタの数や層数なども含む。学習データは、光源情報ごとに学習されており、対応した学習データを使用する。その結果、より精度の高い推定が可能となる。なお、学習の詳細に関しては後述する。

**【0064】**

続いてステップS304において、推定部303は、撮影画像から部分画像を取得する。続いてステップS305において、推定部303は、ステップS303にて決定された学習データに基づいて、部分法線情報を生成する。続いてステップS306において、推定部303は、撮影画像のうち所定の領域（部分画像）の全てに対して法線情報の推定処理が完了したか否かを判定する。所定の領域の全てに対して法線情報の推定が完了した場合、ステップS307へ進む。一方、法線情報の推定が完了していない領域（部分画像）が残っている場合、ステップS304へ戻り、推定部303は、まだ法線情報が推定されていない部分画像（新たな部分画像）を撮影画像から取得する。ステップS307において、推定部303は、法線情報を出力する。

20

**【0065】**

なお、光源からの入射光が被写体の一部にしか届かないなど、撮影画像中の位置に応じて被写体に対する光源環境が大きく異なる場合、推定部303は、ステップ304をステップS302、S303の前に実行することが好ましい。このとき推定部303は、ステップS302、S303において、撮影画像の局所領域に対して光源情報を取得して、対応する学習データなどを取得する。

30

**【0066】**

次に、サーバ305の学習部307により行われる学習データの学習に関して説明する。本実施例において、学習部307は、光源情報ごとに異なる学習データを学習する。学習方法は、図5を参照して実施例1にて説明した方法と同様である。

**【0067】**

まず、異なる光源環境下における複数の撮影画像と法線情報とからなる学習データをシミュレーション（CGレンダリング）によって生成する場合に関して述べる。この場合、光源情報を異なる複数のパターンに設定して法線情報からレンダリング画像を生成し、一対の学習データを取得する。学習部307は、取得した学習データに対してステップS201～S204を実行し、その後、異なる光源情報に対して同様の手順を繰り返す。

40

**【0068】**

また、形状が既知の実被写体を用いて学習データを生成する場合に関して述べる。この場合、既知の形状を有する被写体に対して推定時に用いる光源環境を用意して撮影画像を取得することで、学習データを取得する。被写体の位置情報に応じて学習データを変更する場合、被写体の距離を変化させながら撮影画像を取得すればよい。そして、同じ光源情報の撮影画像ごとにステップS201～S204を実行して学習データを生成する。

**【0069】**

本実施例によれば、反射特性や形状の制約がない一般的な被写体において、撮像画像が

50

ら高精度に法線情報を推定可能な画像処理システムを提供することができる。

【0070】

(その他の実施例)

本発明は、上述の実施例の1以上の機能を実現するプログラムを、ネットワーク又は記憶媒体を介してシステム又は装置に供給し、そのシステム又は装置のコンピュータにおける1つ以上のプロセッサがプログラムを読み出し実行する処理でも実現可能である。また、1以上の機能を実現する回路(例えば、ASIC)によっても実現可能である。

【0071】

各実施例によれば、反射特性や形状の制約がない一般的な被写体に関して、撮像画像から高精度に法線情報を推定することが可能な画像処理方法、画像処理装置、撮像装置、画像処理プログラム、および、記憶媒体を提供することができる。

10

【0072】

以上、本発明の好ましい実施例について説明したが、本発明はこれらの実施例に限定されず、その要旨の範囲内で種々の変形及び変更が可能である。

【符号の説明】

【0073】

102 画像処理部

102b、303 推定部(画像処理部)

103、302 記憶部

301 画像処理装置

20

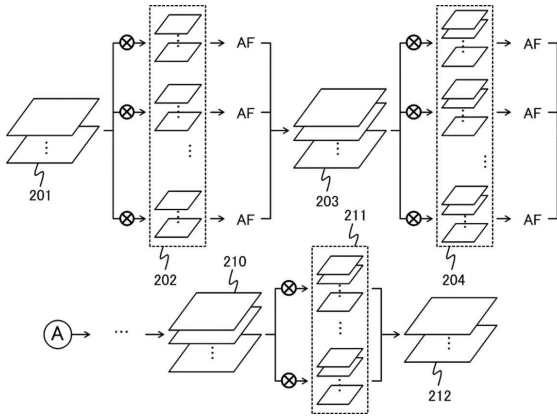
30

40

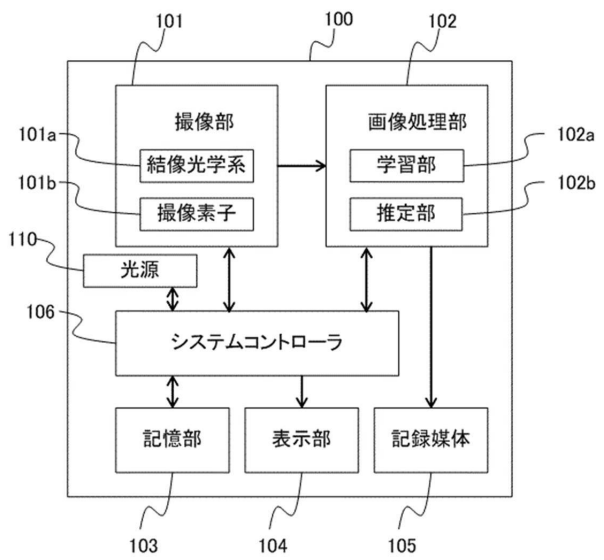
50

【図面】

【図 1】

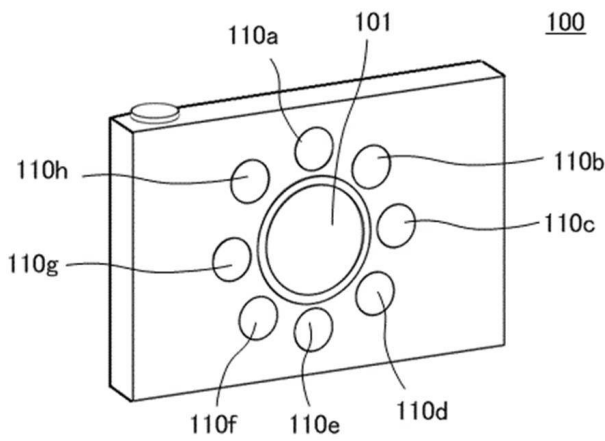


【図 2】

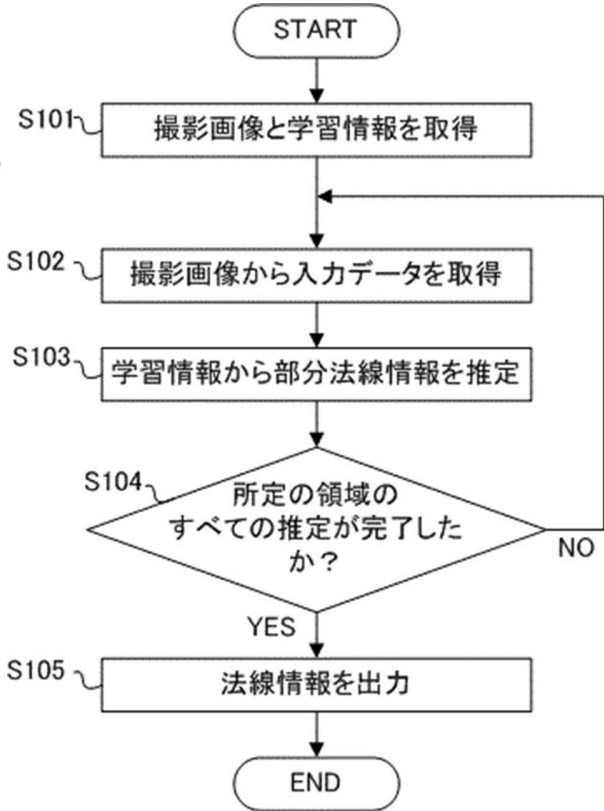


10

【図 3】



【図 4】



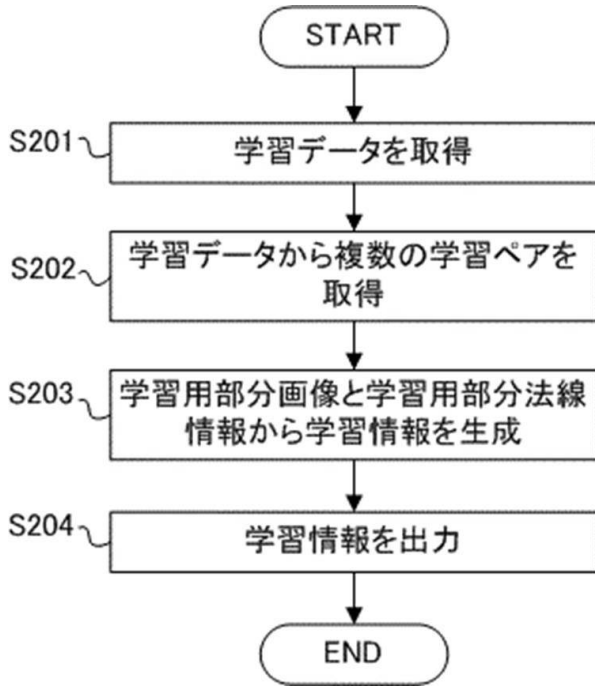
20

30

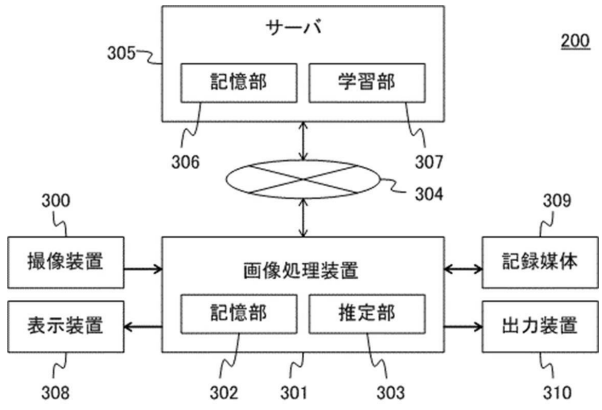
40

50

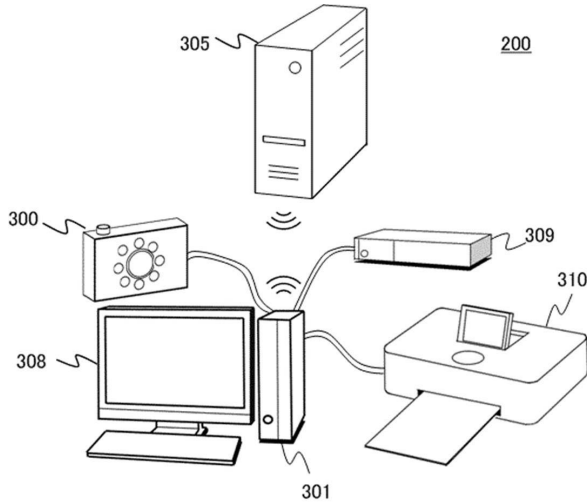
【図5】



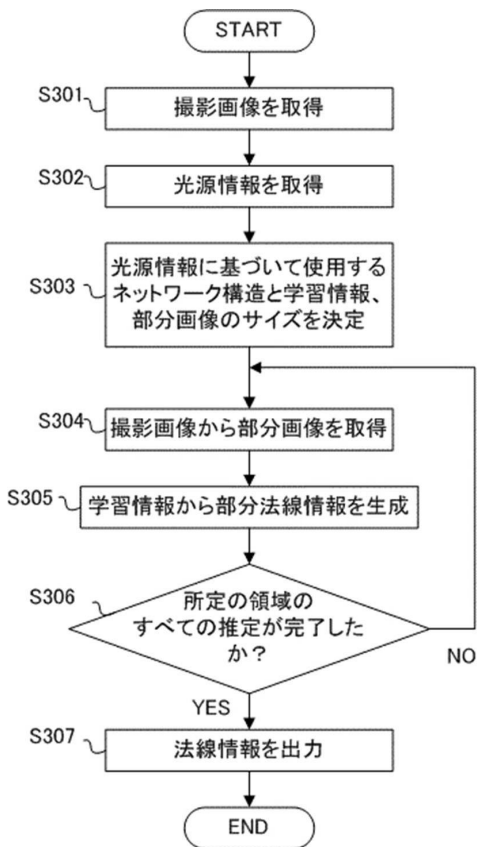
【図6】



【図7】



【図8】



10

20

30

40

50

---

フロントページの続き

- (56)参考文献 特開 2 0 1 7 - 1 6 1 3 7 4 ( J P , A )  
特開 2 0 1 7 - 1 5 0 8 7 8 ( J P , A )  
特開 2 0 1 7 - 1 3 5 5 2 8 ( J P , A )  
特開 2 0 1 7 - 1 3 4 5 6 1 ( J P , A )  
国際公開第 2 0 1 1 / 0 7 0 7 0 8 ( W O , A 1 )  
Silvano Galliani, Konrad Schindler , "Just look at the image: viewpoint-specific surface normal prediction for improved multi-view reconstruction" , 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) , IEEE , 2016年06月27日 , pp.5479-5487 , <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7780960>
- (58)調査した分野 (Int.Cl. , D B 名)  
G 0 6 T 7 / 0 0  
G 0 6 T 7 / 6 0  
G 0 1 B 1 1 / 2 6